

Reasoning with Language Model is Planning with World Model

Year: 2023 | Citations: 796 | Authors: Shibo Hao, Yi Gu, Haodi Ma

Abstract

Large language models (LLMs) have shown remarkable reasoning capabilities, especially when prompted to generate intermediate reasoning steps (e.g., Chain-of-Thought, CoT). However, LLMs can still struggle with problems that are easy for humans, such as generating action plans for executing tasks in a given environment, or performing complex math, logical, and commonsense reasoning. The deficiency stems from the key fact that LLMs lack an internal $\text{\textit{world model}}$ to predict the world $\text{\textit{state}}$ (e.g., environment status, intermediate variable values) and simulate long-term outcomes of actions. This prevents LLMs from performing deliberate planning akin to human brains, which involves exploring alternative reasoning paths, anticipating future states and rewards, and iteratively refining existing reasoning steps. To overcome the limitations, we propose a new LLM reasoning framework, \underline{R} reasoning via \underline{a} planning $\text{\textbf{(RAP)}}$. RAP repurposes the LLM as both a world model and a reasoning agent, and incorporates a principled planning algorithm (based on Monte Carlo Tree Search) for strategic exploration in the vast reasoning space. During reasoning, the LLM (as agent) incrementally builds a reasoning tree under the guidance of the LLM (as world model) and task-specific rewards, and obtains a high-reward reasoning path efficiently with a proper balance between exploration vs. exploitation. We apply RAP to a variety of challenging reasoning problems including plan generation, math reasoning, and logical inference. Empirical results on these tasks demonstrate the superiority of RAP over various strong baselines, including CoT and least-to-most prompting with self-consistency. RAP on LLAMA-33B surpasses CoT on GPT-4 with 33% relative improvement in a plan generation setting.